A Bayesian Approach to Hardness Ratios for non-Bayesians

John E. Davis
<davis@space.mit.edu>

December 18, 2007

1 Introduction

Hardness ratios are often used in situations where there may not be enough counts or spectral resolution to model the spectrum of a source using the traditional forward-folding[Davis(2001)] methodology. In such a case, a statistic formed from the ratio of the counts in one band to another band can serve as a crude characterization of the spectrum. In the low counts regime where this technique might be used, this statistic suffers from significant statistical fluctuations forcing one to consider its confidence limits derived from the statistics underlying probability distribution. The computation of confidence limits for hardness ratios has been addressed by others ([Park et al.(2006), Jin et al.(2006)]) using Bayesian techniques. The present work presents a simple extension of the work by previous authors to a more general class of functions. Although a Bayesian approach was adopted no prior knowledge of Bayesian statistics is assumed on the part of the reader.

2 Definitions

Let C_H denote the number of counts in a "hard" energy band extracted from a spatial region Ω_s, and let C_S denote the counts extracted from a "soft" band. It is assumed that the bands are non-overlapping and that the same extraction region Ω_s was used for both bands. The counts in each of these bands will consist of both source counts and background counts. Let μ_H denote the expected number[1]of source counts in the

[1]The "expected number" is a statistical quantity whose value is not restricted to integers. More precisely, it refers to the expectation value of the of the number of counts with respect to their underlying probability distribution.

hard band for this region. The expected number of hard-band background counts in the region may be estimated from the number of counts extracted from a source-free region Ω_b. Let ν_H denote the expected number of counts in the hard-band of the source-free region, and let α_H be a scaling factor such that the expected number of hard-band background counts in the source region is $\alpha_H \nu_H$. Hence, $\mu_H + \alpha\nu_H$ is the expected number of counts in the region Ω_s for the hard-band. Analogous quantities μ_S, ν_S, and α_S, may be defined for the soft-band. In the following, the scale factors α_H and α_S are assumed to be specified constants, and may be thought of as the ratio of the source to background effective areas for the corresponding bands.

Following [Park et al.(2006)], the so-called "hardness ratio" R is defined by

$$R = \frac{\mu_H}{\mu_S}. \tag{1}$$

In addition to R, [Park et al.(2006)] considered two other functional forms of R: the fractional difference

$$\frac{R-1}{R+1} \tag{2}$$

and the so-called "color" given by $\log_{10}(R)$. In the present work, an arbitrary monotonic function of R is considered, i.e.,

$$\mathcal{H} = f(R). \tag{3}$$

Such a more generic form will allow a hardness-ratio of the form

$$\begin{aligned} \mathcal{H} &= \frac{\mu_H/a_H - \mu_S/a_S}{\mu_H/a_H + \mu_S/a_S} \\ &= \frac{(a_S/a_H)R - 1}{(a_S/a_H)R + 1}, \end{aligned} \tag{4}$$

where, e.g., a_S represents the average effective area for the soft-band and μ_S/a_S is representative of the flux in the band.

Given the *observed* counts C_H, B_H, C_S, and B_S, the goal is to derive a probability distribution for $\mathcal{H}$, and from that its confidence limits may be obtained.

3 Bayes Theorem

Assuming that the source and background counts are Poisson distributed, the probability $\mathcal{P}(C|\mu)$ of obtaining C counts when μ are expected is given by

$$\mathcal{P}(C|\mu) = \frac{\mu^C}{C!} e^{-\mu}. \tag{5}$$

Note that the observed number of counts C is integer-valued, whereas the expected number μ is a real-valued quantity. The probability $\mathcal{P}(C|\mu)$ should not be interpreted as the probability density for the expected number of counts μ given the observed number C. The latter quantity is denoted as $\mathcal{P}(\mu|C)$, and may be found as follows: The *joint* probability for obtaining exactly C counts *and* for the expected number of counts to lie between μ and $\mu + \mathrm{d}\mu$ is denoted as $\mathcal{P}(\mu, C)\mathrm{d}\mu$. It is related to the conditional probabilities via

$$\mathcal{P}(\mu, C) = \mathcal{P}(\mu|C)\mathcal{P}(C) = \mathcal{P}(C|\mu)\mathcal{P}(\mu). \tag{6}$$

From the above it follows that

$$\mathcal{P}(\mu|C) = \frac{\mathcal{P}(C|\mu)\mathcal{P}(\mu)}{\mathcal{P}(C)} \propto \mathcal{P}(C|\mu)\mathcal{P}(\mu), \tag{7}$$

which is known as Bayes' Theorem. By itself, it is not terribly useful without specifying a value for $\mathcal{P}(\mu)$. Note that $\mathcal{P}(\mu)$ is related to the joint probability distribution via

$$\mathcal{P}(\mu) = \sum_C \mathcal{P}(\mu, C). \tag{8}$$

This is also an empty statement without knowing anything about the joint distribution. Hence, methodologies utilizing Bayes's Theorem require some prior knowledge of $\mathcal{P}(\mu)$, and for this reason statisticians calls refer to $\mathcal{P}(\mu)$ as simply the *prior*. For now, $\mathcal{P}(\mu)$ will simply be regarded as a function to be specified later.

4 Probability Distribution for the Hardness Ratio

The probability density for $\mathcal{H}$ given the observed counts is given by

$$\mathcal{P}(\mathcal{H}) = \int_0^\infty \mathrm{d}\mu_H \int_0^\infty \mathrm{d}\mu_S \; \mathcal{P}(\mu_H|C_H, B_H)\mathcal{P}(\mu_S|C_S, B_S)\delta(H - f(\frac{\mu_H}{\mu_S})) \tag{9}$$

Note that the delta function constrains the ratio μ_H/μ_S to a fixed value R satisfying $\mathcal{H} = f(R)$. Since $f(R)$ is assumed to be monotonic, the delta function may be written in the form

$$\delta(\mathcal{H} - f(\frac{\mu_H}{\mu_S})) = \frac{\delta(\mu_H - \mu_H^0)}{|\partial f/\partial \mu_H|}, \tag{10}$$

where $\mu_H^0 = \mu_S f^{-1}(\mathcal{H}) = R\mu_S$ and $\partial f/\partial \mu_H = f'(R)/\mu_S$. Substituting this expression for the delta function into equation (9) and integrating over μ_H yields

$$\mathcal{P}(\mathcal{H}) = \frac{1}{f'(R)} \int_0^\infty \mathrm{d}\mu_S \; \mu_S \mathcal{P}(R\mu_S|C_H, B_H)\mathcal{P}(\mu_S|C_S, B_S). \tag{11}$$

The probabilities $\mathcal{P}(\mu_H|C_H, B_H)$ and $\mathcal{P}(\mu_S|C_S, B_S)$ may be obtained as follows. For clarity, consider one of the bands, say the soft band, and drop the subscript denoting the band. Then the joint probability $\mathcal{P}(\mu,\nu,C,B)$ may be factored as both

$$\mathcal{P}(\mu,\nu,C,B) = \mathcal{P}(\mu,\nu|C,B)\mathcal{P}(C,B) \tag{12}$$

and

$$\mathcal{P}(\mu,\nu,C,B) = \mathcal{P}(C|\mu,\nu,B)\mathcal{P}(\mu,\nu,B). \tag{13}$$

The fact that B and C were extracted from independent regions, means that the value of C cannot depend upon B, nor can the expected number of source counts μ depend upon B. Hence,

$$P(C|\mu,\nu,B) = \mathcal{P}(C|\mu,\nu) \tag{14}$$

and

$$\begin{aligned}\mathcal{P}(\mu,\nu,B) &= \mathcal{P}(\mu)\mathcal{P}(\nu,B)\\ &= \mathcal{P}(\mu)\mathcal{P}(B|\nu)\mathcal{P}(\nu)\end{aligned} \tag{15}$$

From these expressions it follows that

$$\mathcal{P}(\mu,\nu|C,B) = \frac{\mathcal{P}(C|\mu,\nu)\mathcal{P}(B|\nu)\mathcal{P}(\mu)\mathcal{P}(\nu)}{\mathcal{P}(C)\mathcal{P}(B)}, \tag{16}$$

and

$$\begin{aligned}\mathcal{P}(\mu|C,B) &= \int_0^\infty \mathrm{d}\nu P(\mu,\nu|C,B)\\ &\propto \mathcal{P}(\mu)\int_0^\infty \mathrm{d}\nu\ \mathcal{P}(C|\mu,\nu)\mathcal{P}(B|\nu)\mathcal{P}(\nu),\end{aligned} \tag{17}$$

where the proportionality constant depends upon C and B, and may be obtained by demanding that $\mathcal{P}(\mu|C,B)$ is normalized according to

$$1 = \int_0^\infty \mathrm{d}\mu\ \mathcal{P}(\mu|C,B). \tag{18}$$

From the assumption of Poisson statistics,

$$\mathcal{P}(C|\mu,\nu) = \frac{(\mu+\alpha\nu)^C}{C!}e^{-(\mu+\alpha\nu)}, \tag{19}$$

and

$$\mathcal{P}(B|\nu) = \frac{\nu^B}{B!}e^{-\nu}. \tag{20}$$

From these expressions, it follows that

$$\begin{aligned}\mathcal{P}(\mu|C,B) &\propto \frac{\mathcal{P}(\mu)e^{-\mu}}{C!B!}\int_0^\infty \mathrm{d}\nu\ \mathcal{P}(\nu)(\mu+\alpha\nu)^C\nu^B e^{-(\alpha+1)\nu}\\ &\propto \frac{\mu^C}{C!}e^{-\mu}\mathcal{P}(\mu)\sum_{j=0}^{C}\binom{C}{j}(\frac{\alpha}{\mu})^j\frac{1}{B!}\int_0^\infty \mathrm{d}\nu\ \mathcal{P}(\nu)\nu^{B+j}e^{-(\alpha+1)\nu},\end{aligned} \tag{21}$$

where

$$\binom{C}{j} = \frac{C!}{j!(C-j)!} \tag{22}$$

denotes the binomial coefficient. The above may be expressed in a slightly simpler form as

$$\mathcal{P}(\mu|C,B) \propto \mathcal{P}(\mu)\frac{\mu^C}{C!}e^{-\mu}\sum_{j=0}^{C}\binom{C}{j}(\frac{\alpha}{\mu})^j\frac{I_j(B,\alpha)}{B!}, \tag{23}$$

where

$$I_j(B,\alpha) = \int_0^\infty \mathrm{d}\nu\ \mathcal{P}(\nu)\nu^{B+j}e^{-(\alpha+1)\nu}. \tag{24}$$

In order to compute $I_j(B,\alpha)$, the probability distribution $\mathcal{P}(\nu)$ must be specified. In the absence of any prior knowledge of this function, the so-called *uniform prior*, which assumes that $\mathcal{P}(\nu)$ is independent of ν is usually adopted. This choice produces

$$I_j(B,\alpha) = \frac{(B+j)!}{(1+\alpha)^{B+j+1}} \quad \text{(uniform prior).} \tag{25}$$

Alternatively, a specific parameterization of $\mathcal{P}(\nu)$ such as $\nu^m e^{-\lambda\nu}$ could be used. The use of this particular form has the effect of shifting B to $B' = B + m$, and α to $\alpha' = \alpha+\lambda$. This parameterization of $\mathcal{P}(\nu)$ also has the property that since $\mathcal{P}(B|\nu)$ is Poisson distributed, then $\mathcal{P}(\nu|B)$ computed via equation (7) also has the same functional form as $\mathcal{P}(\nu)$. Statisticians use the term *conjugate prior* for parameterizations that have this property. Hence,

$$I_j(B,\alpha) = \frac{\Gamma(B'+j+1)}{(1+\alpha')^{B'+j+1}} \quad \text{(conjugate prior).} \tag{26}$$

As an example of the a conjugate prior, suppose that an independent measurement of the Poisson-distributed background yielded B_1 counts. Using this measurement and Bayes's Theorem with a uniform prior, it is easy to show that

$$\mathcal{P}(\nu|B_1) = \frac{\nu^{B_1}}{B_1!}e^{-\nu}, \tag{27}$$

which has the form of a conjugate prior for the Poisson distribution. Hence, if this prior were used then $B' = B + B_1$ and $\alpha' = \alpha + 1$. It is also interesting to note that, for this prior, the expected value for ν is given by

$$\bar{\nu} = \int_0^\infty \nu\mathcal{P}(\nu|B_1) = B_1 + 1 \tag{28}$$

Substituting equation (23) into equation (11) and simplifying yields

$$\mathcal{P}(\mathcal{H}) \propto \frac{1}{f'(R)} \sum_{j=0}^{C_H} \sum_{k=0}^{C_S} \left\{ \binom{C_H}{j} \binom{C_S}{k} \frac{\alpha_H^j \alpha_S^k}{C_H! C_S!} R^{C_H - j} \frac{I_j(B_H, \alpha_H)}{B_H!} \frac{I_k(B_S, \alpha_S)}{B_S!} \right.$$
$$\left. \times \int_0^\infty \mathrm{d}\mu_S \; \mathcal{P}(\mu_S) \mathcal{P}(R\mu_S) \mu_S^{C_H + C_S + 1 - j - k} e^{-\mu_S (R+1)} \right\} \tag{29}$$

With the adoption of uniform priors for $\mathcal{P}(\mu_S)$ and $\mathcal{P}(\mu_H)$, the above equation may be simplified to

$$\mathcal{P}(\mathcal{H}) \propto \frac{1}{f'(R)} \sum_{j=0}^{C_H} \sum_{k=0}^{C_S} \left\{ \binom{C_H}{j} \binom{C_S}{k} \frac{(j+k+1)!}{C_H! C_S! B_H! B_S!} \right.$$
$$\left. \times I_{C_H - j}(B_H, \alpha_H) I_{C_S - k}(B_S, \alpha_S) \alpha_H^{C_H - j} \alpha_S^{C_S - j} \frac{R^j}{(R+1)^{j+k+2}} \right\} \tag{30}$$

For the specific case of conjugate priors for the expected background counts, equation (26) may be substituted into the above equation and simplified to produce

$$\mathcal{P}(\mathcal{H}) \propto \frac{1}{f'(R)} \sum_{j=0}^{C_H} \sum_{k=0}^{C_S} A_{jk} \left(\frac{\alpha_H}{1 + \alpha'_H}\right)^{C_H - j} \left(\frac{\alpha_S}{1 + \alpha'_S}\right)^{C_S - k} \frac{R^j}{(R+1)^{j+k+2}}, \tag{31}$$

where

$$A_{jk} = \binom{C_H}{j} \frac{\Gamma(1 + B'_H + C_H - j)}{C_H! B_H!} \binom{C_S}{k} \frac{\Gamma(1 + B'_S + C_S - k)}{C_S! B_S!} (j+k+1)!. \tag{32}$$

5 A Maximum-Likelihood Estimate

Let $\lambda_H = \mu_H + \alpha_H \nu_H$ denote the expected number of counts in the hard-band source region. A maximum-likelihood estimate of this quantity is C, the observed number of counts in the region. Similarly, the maximum-likelihood estimate of ν_H is B_H. Then it follows that a maximum-likelihood estimate of $\mathcal{P}(\mu_H | C_H, B_H)$ is given by

$$P(\mu_H | C_H, B_H) = \frac{\mu_H^{C_H - \alpha_H B_H}}{\Gamma(1 + C_H - \alpha_H B_H)} e^{-\mu_H}, \tag{33}$$

where the gamma function has been used instead of the factorial since $C_H - \alpha B_H$ is not necessarily an integer. Inserting this expression and the analogous one for the soft-band into equation (11) and integrating yields

$$\mathcal{P}(\mathcal{H}) = \frac{\Gamma(2 + C_H - \alpha_H B_H + C_S - \alpha_S B_S)}{\Gamma(1 + C_H - \alpha_H B_H) \Gamma(1 + C_S - \alpha_S B_S)}$$
$$\times \frac{1}{f'(R)} \frac{R^{C_H - \alpha_H B_H}}{(R+1)^{C_H - \alpha_H B_H + C_S - \alpha_S B_S + 2}}. \tag{34}$$

Another way of seeing this is to look for the dominant term of equation (31), which is the term that contributes the largest contribution to the integrated probability. To this end, assume $\mathcal{H} = f(R) = R$ and integrate equation (31) over R to obtain

$$\int_0^\infty \mathrm{d}R\, \mathcal{P}(R) \propto \Big[\sum_{j=0}^{C_H} \frac{\Gamma(1+B'_H+C_H-j)}{\Gamma(1+C_H-j)} \big(\frac{\alpha_H}{1+\alpha'_H}\big)^{C_H-j}\Big] \times \Big[\sum_{k=0}^{C_S} \frac{\Gamma(1+B'_S+C_S-k)}{\Gamma(1+C_S-k)} \big(\frac{\alpha_S}{1+\alpha'_S}\big)^{C_S-k}\Big]. \tag{35}$$

Let the $T(j)$ denote the jth term in the sum over j, i.e.,

$$T(j) = \frac{\Gamma(1+B'_H+C_H-j)}{\Gamma(1+C_H-j)} \big(\frac{\alpha_H}{1+\alpha'_H}\big)^{C_H-j}, \tag{36}$$

and regard it as a continuous function of j. This function will take on its maximum value when j satisfies $\mathrm{d}T/\mathrm{d}j = 0$. By assuming the asymptotic form

$$\log\Gamma(1+x) \sim x\log x - x, \tag{37}$$

it easily follows that

$$j = \frac{(C_H - \alpha_H B'_H) + C_H(\alpha'_H - \alpha_H)}{1 + (\alpha'_H - \alpha)}. \tag{38}$$

Since this equation was derived assuming the asymptotic form of $\Gamma(x)$, it is expected to hold when $C_H - j$ is sufficiently large. Recall that for a uniform prior, $\alpha_H = \alpha'_H$ and $B_H = B'_H$, and as a result $j = C_H - \alpha_H B_H$. Hence, the maximum likelihood estimate given by equation (34) corresponds to the dominant term of the Bayesian result. Similarly, for the background prior corresponding to equation (27) with $\alpha'_H = \alpha_H + 1$ and $B'_H = B_H + B_1$, one obtains $j = C_H - \alpha_H(B_H + B_1)/2$, whose interpretation is also consistent with the maximum-likelihood result with $(B_H + B_1)/2$ viewed as an improved estimate of the background.

6 Conclusion

Equation (31) is the main result of this work. It assumes uniform priors[2]for the expected source counts in the hard and soft bands, but allows conjugate priors to be used for the expected background counts. The equation describing the case involving uniform priors for the background counts may be obtained by simply dropping the primes from equation (31). As it describes the probability function for hardness ratio defined by an arbitrary function of $f(R)$, it should be regarded as an extension of the work by [Park et al.(2006)].

[2]This restriction is easily lifted by assuming conjugate priors for the expected source counts. This minor complication would manifest itself as additional primed quantities of an equation resembling equation (31).

Acknowledgements

I am grateful to Mike Nowak for the useful insights into Bayesian methods and for providing me with code that enabled me to make detailed comparisons with his work.

References

[Davis(2001)] Davis, J.E., 2001, ApJ, 548, 1010

[Park et al.(2006)] Park, T., Kashyap, V.L., Siemiginowska, A., van Dyk, D., Heinke, C., Wargelin, B.J., 2006, ApJ, 652, 610

[Jin et al.(2006)] Jin, Y.K., Zhang, S.N., Wu, J.F., 2006, ApJ 653, 1566 Wargelin, B.J., 2006, ApJ, 652, 610